Lateral transfers of large DNA fragments spread functional genes among grasses

Luke T. Dunning[a], Jill K. Olofsson[a], Christian Parisod[b], Rimjhim Roy Choudhury[b], Jose J. Moreno-Villena[a,1], Yang Yang[c], Jacqueline Dionora[d], W. Paul Quick[a,d], Minkyu Park[e], Jeffrey L. Bennetzen[e], Guillaume Besnard[f], Patrik Nosil[a], Colin P. Osborne[a], and Pascal-Antoine Christin[a,2]

[a]Animal and Plant Sciences, University of Sheffield, Western Bank, S10 2TN Sheffield, United Kingdom; [b]Institute of Plant Sciences, University of Bern, 3013 Bern, Switzerland; [c]Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650204 Yunnan, China; [d]Systems Physiology Cluster, International Rice Research Institute, 1301 Metro Manila, Philippines; [e]Department of Genetics, University of Georgia, Athens, GA 30602; and [f]Laboratoire Évolution & Diversité Biologique (EDB UMR5174), CNRS, Institut de Recherche pour le Développement, F-31062 Toulouse, France

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved January 17, 2019 (received for review June 11, 2018)

A fundamental tenet of multicellular eukaryotic evolution is that vertical inheritance is paramount, with natural selection acting on genetic variants transferred from parents to offspring. This lineal process means that an organism's adaptive potential can be restricted by its evolutionary history, the amount of standing genetic variation, and its mutation rate. Lateral gene transfer (LGT) theoretically provides a mechanism to bypass many of these limitations, but the evolutionary importance and frequency of this process in multicellular eukaryotes, such as plants, remains debated. We address this issue by assembling a chromosome-level genome for the grass *Alloteropsis semialata*, a species surmised to exhibit two LGTs, and screen it for other grass-to-grass LGTs using genomic data from 146 other grass species. Through stringent phylogenomic analyses, we discovered 57 additional LGTs in the *A. semialata* nuclear genome, involving at least nine different donor species. The LGTs are clustered in 23 laterally acquired genomic fragments that are up to 170 kb long and have accumulated during the diversification of *Alloteropsis*. The majority of the 59 LGTs in *A. semialata* are expressed, and we show that they have added functions to the recipient genome. Functional LGTs were further detected in the genomes of five other grass species, demonstrating that this process is likely widespread in this globally important group of plants. LGT therefore appears to represent a potent evolutionary force capable of spreading functional genes among distantly related grass species.

adaptation | genome | Poaceae | horizontal gene transfer | phylogenetics

During evolution, organisms adapt to new or changing environments as a result of natural selection acting on genetic variation. In multicellular eukaryotes, this process is traditionally considered to concern mutations transferred from parents to offspring. The possibility of any given organism evolving novel traits can therefore be constrained by the genetic variants existing within an interbreeding population or species, and the rate of new mutations (1). Therefore, a novel trait can take protracted evolutionary periods to develop, with incremental modifications per generation. Furthermore, divergent evolutionary histories may mean that particular traits are restricted to lineages that possess the appropriate genetic precursors (2). The transfer of genes among distantly related species can theoretically allow organisms to bypass these limitations. However, the frequency of this phenomenon, and therefore its importance for the evolutionary diversification of multicellular eukaryotes, remains unclear (3–10).

Lateral gene transfer (LGT) is the movement of genetic material between organisms belonging to distinct groups of interbreeding individuals, and therefore involves mechanisms other than classical sexual reproduction. Its pervasiveness is well documented in prokaryotes, where it can rapidly spread adaptive traits such as antibiotic resistance among distantly related taxa (11). Reports of LGTs have also been accumulating for multiple groups of eukaryotes, frequently involving unicellular recipients or donors (e.g., refs. 12–15). While LGTs have been less commonly reported among multicellular eukaryotes, convincing cases exist where genes of adaptive significance have been transferred (e.g., refs. 3, 16, and 17). Among plants, most known LGTs concern mitochondrial genes (18–21) and/or parasitic interactions (22–30), with only a few nonparasitic plant-to-plant LGT of nuclear genes yet recorded (31–34). Genome scans suggest that transposable elements (TEs) are frequently transferred among plants (35), but similar searches of laterally acquired coding genes are needed to assess the frequency and functional significance of nuclear LGT among nonparasitic plants.

To determine the importance of LGTs for functional diversification in a group of nonparasitic plants, we quantified the prevalence, retention through time, and functional significance of LGT in the grass *Alloteropsis semialata* (tribe Paniceae in subfamily Panicoideae). This species is distributed throughout the paleotropics and exhibits geographic variation in the presence of two genes encoding key C_4 photosynthetic enzymes that were laterally acquired from distantly related grass species (32, 36). The donor species diverged from *A. semialata* over 20 My

Significance

In multicellular organisms, exchange of genetic information occurs mainly among individuals belonging to the same species through sexual reproduction. Lateral gene transfer between distantly related taxa has been demonstrated in some cases, but its frequency and evolutionary importance have been controversial. By comparing genomes of many grasses, we show that large blocks of DNA containing functional genes are laterally passed among distantly related species. Some of these genes are then used by the recipient species, expanding their genetic toolkit. The spread of functional genes across grasses that have developed distinct physiological and ecological adaptations may therefore represent a significant evolutionary driving force in this globally important group of plants.

Author contributions: L.T.D., P.N., C.P.O., and P.-A.C. designed research; L.T.D. and J.J.M.-V. performed research; L.T.D., J.K.O., Y.Y., J.D., W.P.Q., M.P., J.L.B., G.B., C.P.O., and P.-A.C. contributed new reagents/analytic tools; L.T.D., J.K.O., C.P., and R.R.C. analyzed data; and L.T.D. and P.-A.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All raw DNA sequencing data (PacBio and Illumina reads) generated as part of this study have been deposited with National Center for Biotechnology Information under BioProject PRJNA481434 (Sequence Read Archive accession nos. SRR7528994–SRR7529026). The *A. semialata* reference genome has been deposited in GenBank under the accession QPGU00000000 (the version described in this paper is QPGU01000000).

[1]Present address: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511.

[2]To whom correspondence should be addressed. Email: p.christin@sheffield.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810031116/-/DCSupplemental.

Published online February 20, 2019.

ago (32), more than ample time for the complete turnover of intergenic regions (37–39). Classical introgression involving chromosome pairing and recombination is therefore unlikely (32). There are no known parasitic grasses (40), and *A. semialata* must therefore have received its additional genetic information via a different transfer mechanism than those reported for symbiotic partners (1). One of the two LGTs previously detected in *A. semialata* appears to be restricted to Australia (36), enabling comparisons between closely related individuals with and without it. The identification of the putative donor (32), coupled with such recent LGTs, provides a tractable system to evaluate the evolutionary significance of grass-to-grass LGT.

In this present work, we generate and assemble a chromosome-level genome for an Australian individual of *A. semialata* known to contain one LGT received from the distantly related Panicoideae grass *Themeda* (tribe Andropogoneae), and one received from a more closely related grass species belonging to a different subtribe within the Paniceae (*Cenchrus* sp. in the Cenchrinae) (32, 36). Using genomic data from another 146 grasses, including members of the groups known to have donated genes to certain *Alloteropsis* populations (32), we then adopt stringent phylogenomic approaches to first identify all unambiguous LGTs in the reference genome of *A. semialata* and to determine for each of them the putative donor. The genomic data are then used to determine the size and number of DNA blocks that were laterally acquired in *Alloteropsis*, as well as the transposable elements that appear to have accompanied them. We then use genomic data for conspecifics and congeners of the reference genome to determine when the genes were acquired during the diversification of *Alloteropsis*, testing the hypothesis that LGTs have gradually accumulated in the genome. Finally, gene expression data are used to determine whether the LGTs are transcribed and test the hypothesis that their acquisition added functional diversity to the recipient genome.

Results

Assembly and Annotation of a Chromosome-Level Reference Genome. We sequenced and assembled a 0.75-Gb chromosome-level reference genome for a single Australian individual of *A. semialata* (*SI Appendix*, *Supplementary Methods* and Fig. S1). In total, 97.5% of the genome assembly was contained in nine scaffolds, which corresponds to the expected number of chromosomes for this individual ($2n = 18$) (41). Synteny is well conserved between the genome of *A. semialata* and those of other Panicoideae grasses, such as the Cenchrinae *Setaria italica* (42) and, to a lesser extent, the Andropogoneae *Sorghum bicolor* (43) (*SI Appendix*, Fig. S2); this is as expected, given high overall synteny among grasses (44, 45).

The genome contains 22,043 high-confidence annotated protein-coding genes, with BLAST matches in both SwissProt and at least one of the two other Panicoideae genomes (*S. italica* or *S. bicolor*). The gene density was plotted across the genome in 1-Mb windows. Each chromosome has a region of reduced gene density, and these regions are assumed to correspond to centromeres (*SI Appendix*, Fig. S3*A*). All putative centromeres are located roughly in the middle of the chromosome, apart from chromosomes 7 and 9 which appear acrocentric (*SI Appendix*, Fig. S3*A*).

Phylogenomics Identifies Multiple LGTs in *A. semialata*. We adopted a tiered approach to determine the proportion of the 22,043 high-confidence annotated genes within the genome that were laterally acquired from other plant species (Fig. 1 and *SI Appendix*, *Supplementary Methods*). By specifically focusing on plant-to-plant transfers, we limit the risks of contamination and false positives associated with LGT studies involving microorganisms (4, 6). We first performed BLAST searches against angiosperm genomes to determine whether any *A. semialata* gene is more similar to a nongrass angiosperm than to other grasses. No such evidence for gene transfer involving a nongrass angiosperm donor was found, and we therefore focused our searches on grass-to-grass LGTs. Our strategy used existing and novel genomic resources for 147 grass species in a pipeline combining similarity analyses with phylogenetic validation (Fig. 1). The analytical approach is analogous to previous scans for LGT (e.g., refs. 29 and 32), but extra validation steps were made possible by the availability of additional genomic information (Fig. 1). We first used a read mapping strategy for a selection of species with high-coverage data ($n = 20$; mean = 42.63 Gb; SD = 5.59 Gb) to identify all high-confidence genes in the *A. semialata* reference genome with a higher percentage identity to one of 17 potential donors [*Themeda triandra* sequenced here and 16 species from National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA); *SI Appendix*, *Supplementary Methods* and Dataset S1] than to the three conspecifics or congeners. This initial genome scan identified 1,148 LGT candidates (5.21% of high-confidence protein-coding genes), although these likely contain many false positives, for instance caused by gene losses in the close relatives. We therefore subjected all 1,148 LGT candidates to phylogenetic investigation, using two successive steps, which first

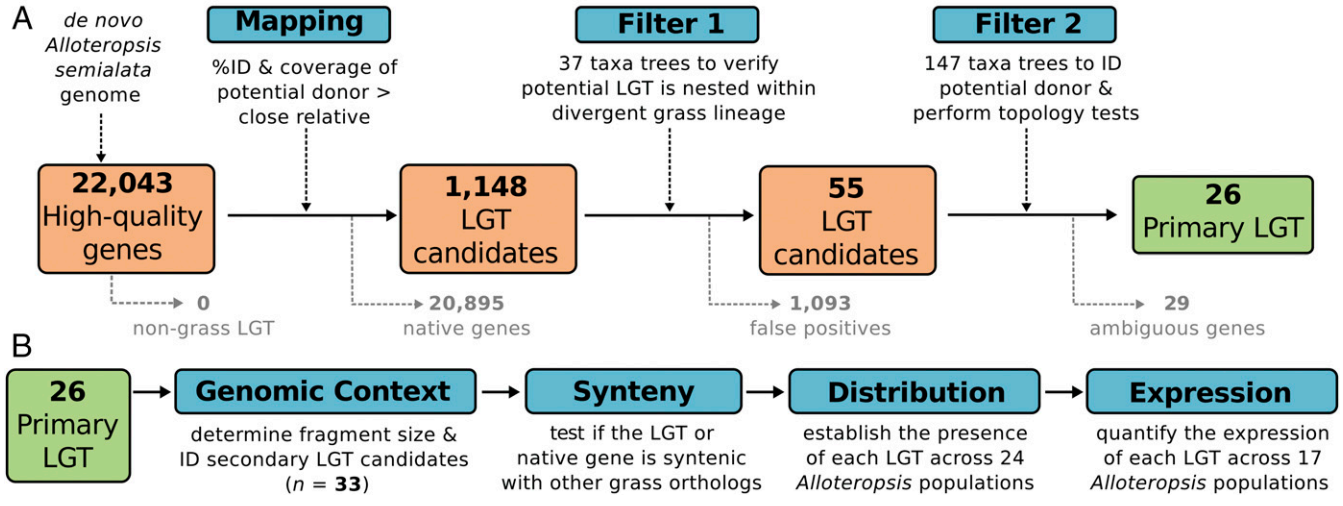

Fig. 1. Overview of the analytical pipeline. (*A*) For each of the steps used to identify lateral gene transfers in the reference genome of *A. semialata*, the number of candidates retained/discarded is indicated. (*B*) The purpose of each set of analyses conducted on the unambiguous LGTs is indicated.

identified and subsequently validated well-supported LGT candidates (Fig. 1).

First, sequences from the *A. semialata* genome were compared with those of 16 completely sequenced grass genomes and 20 grass transcriptomes (*SI Appendix*, *Supplementary Methods* and Dataset S1). A coalescence multigene "species" phylogenetic tree was inferred for these 36 taxa and *A. semialata* using 200 universal single-copy orthologs identified using BUSCO (46). The 37-taxa species tree was generally well resolved, and overall congruent with previous analyses (Fig. 2) (47, 48). While there is some gene tree discordance, particularly for the nodes at the base of the Paniceae, there are well-supported groups spread throughout the phylogenetic tree that are recovered by a majority of the nuclear markers (Fig. 2 and *SI Appendix*, Table S1). Genes of *A. semialata* positioned within any of these well-supported clades should therefore be considered strong LGT candidates. For each of these 1,148 LGT candidates, we then inferred a maximum likelihood phylogenetic tree using the same 37 taxa. If there was support (>50% bootstrap replicates) for the LGT candidate being nested within (not just sister to) one of the groups retrieved by the coalescence species tree, it was retained for further analysis. We further retained sequences of *A. semialata* sister to groups represented by a single sequence, such as Melinidinae, and those outside of the core Panicoideae not assigned to a well-supported clade (i.e., *Stipagrostis hirtigluma*, *Danthonia californica*, or *Chasmanthium latifolium*; Fig. 2) if their combined sister group was as expected based on the species tree (*SI Appendix*, *Supplementary Methods*). The phylogenetic analyses and topological filtering retained 55 LGT candidates (Fig. 1 and *SI Appendix*, Table S1).

Second, we validated these 55 candidates by inferring gene trees with dense phylogenetic sampling using sequence information extracted from 238 genome and transcriptome datasets for 147 grass species (including *A. semialata*; Fig. 3, *SI Appendix*, *Supplementary Methods*, and Dataset S2). To allow comparison with the candidate LGT gene tree, a coalescence species tree was inferred for these 147 species with the same 200 BUSCO markers as used above (*SI Appendix*, *Supplementary Methods* and Fig. S4). The topology of the species tree recovered the main taxonomic groups that have been identified in the reduced phylogenetic tree (Fig. 2) and previous published grass phylogenetic trees based on plastids or a few nuclear markers (47, 48). The 55 candidate LGT gene trees were manually inspected to verify that (*i*) the positioning of *A. semialata* sequences within a group of distant relatives was well supported (>70% bootstrap values), and (*ii*) the number of species represented by different paralogs (identified by comparing the gene and species trees) was sufficient to draw conclusions regarding phylogenetic relationships. In total, five genes were discarded because fewer than three species were represented outside of *Alloteropsis* and the putative donor clades, or in the donor clade. A further 14 genes were discarded because they inferred relationships that strongly differed from the species tree and were deemed phylogenetically unreliable. In two cases, the *A. semialata* gene was not nested in a distant clade with the denser sampling, and in six phylogenetic trees, the nesting was not

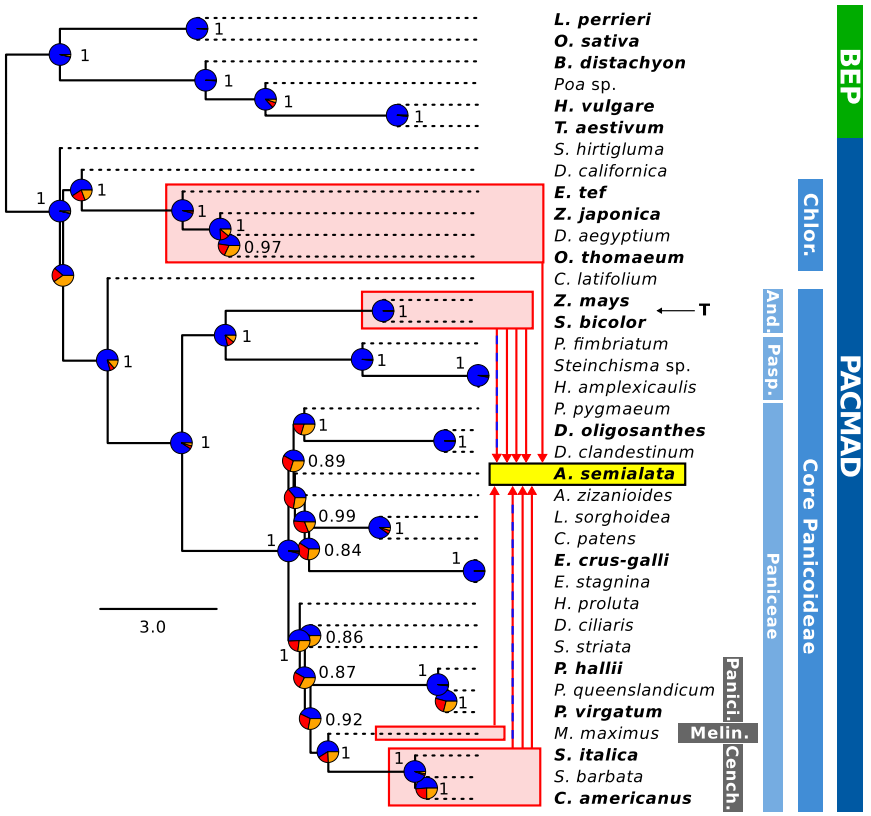

Fig. 2. Multigene coalescence species tree. The relationships are based on 200 single-copy genes extracted from complete genomes (bold species names) and transcriptomes of 37 grass species. The pie charts show the proportion of quartets supporting the species tree topology (blue) and the two alternative topologies (red and orange, respectively). Posterior probabilities supporting values ≥0.50 are shown near nodes, and branch lengths are in coalescent units, with null terminal branches and dashed lines connecting to species names. The main clades of grasses are delimited on the *Right*. The position of the *A. semialata* reference genome is highlighted in yellow, the groups of donors in red, and the clade that contains *Themeda* is indicated. Red arrows represent the transfers of fragments from each identified donor into *A. semialata*, with those reported before (32) indicated with blue dashes. And., Andropogoneae; Cench., Cenchrinae; Chlor., Chloridoideae; Melin., Melinidinae; Panici., Panicinae; and Pasp., Paspaleae.

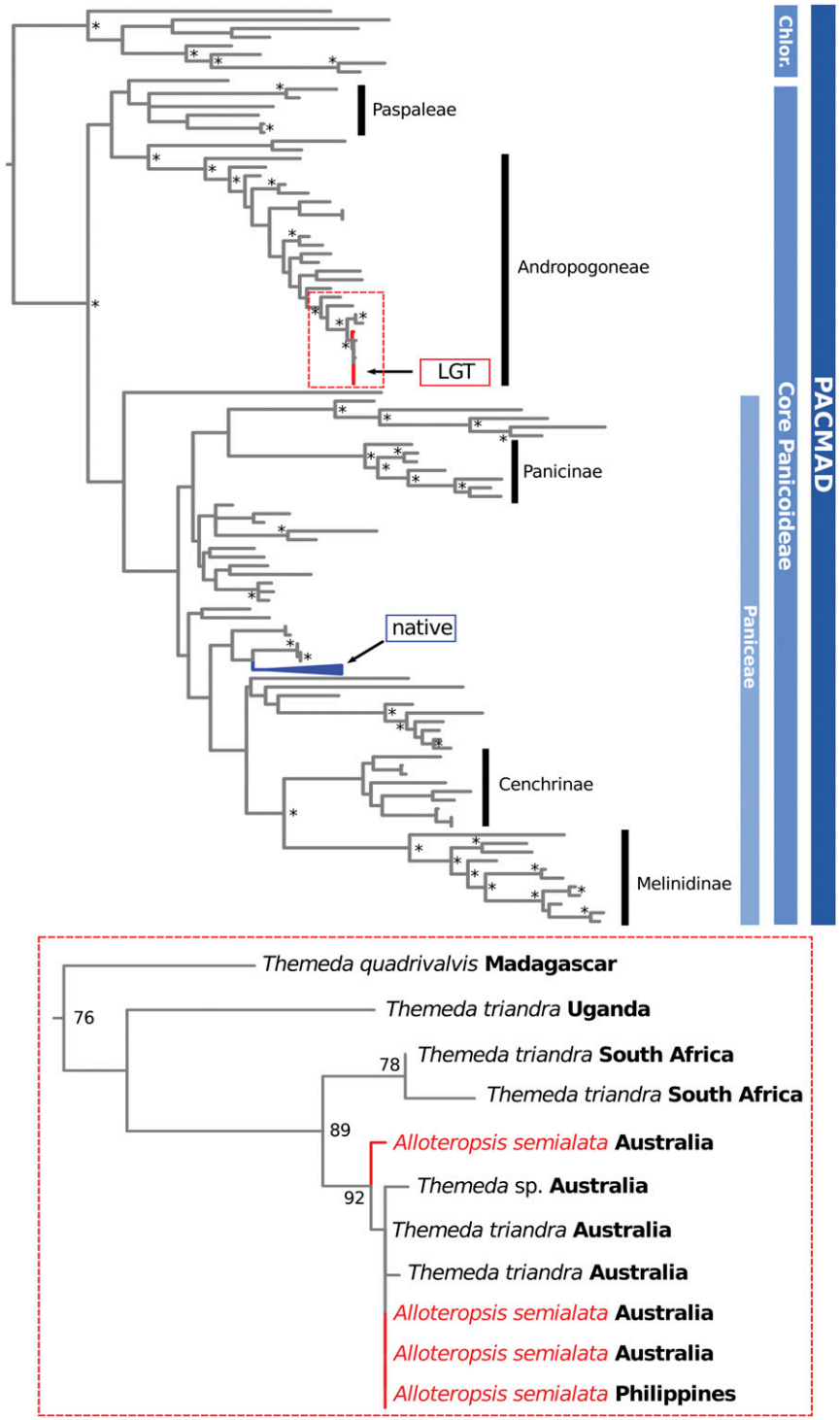

Fig. 3. Phylogenetic evidence for LGT in the reference *A. semialata* genome. The gene ASEM_AUS1_12633 from *A. semialata* was laterally acquired from an Australian *T. triandra*. The maximum likelihood phylogeny inferred from third positions of codons (Dataset S3) is shown, with the LGT in *A. semialata* in red and the native orthologs in blue. The region of the phylogeny containing the LGT [red dashed rectangle (*Upper*) is expanded (*Lower*)]. Bootstrap support values ≥75% are shown (*Lower*) or denoted as asterisks (*Upper*), and the main taxonomic groups are delimited on the *Right* as in Fig. 2. Chlor., Chloridoideae.

supported by bootstrap values above 70%. The remaining 28 candidates were subjected to further validation to rule out alternative scenarios.

We first demonstrated phylogenetic support for the LGT using approximately unbiased (AU) tests, which confirmed that the topology inferring a LGT was significantly better than a topology indicating no LGT in all but two cases (Bonferroni-corrected P value <0.05; *SI Appendix*, Table S2). Second, phylogenetic biases due to adaptive evolution were ruled out by showing that phylogenetic trees based on third codon positions supported the same groupings for all remaining 26 candidates (Dataset S3). Third, unrecognized paralogies were excluded by demonstrating that genes

from *S. italica* and *S. bicolor* are generally syntenic with the copy of *A. semialata* positioned as expected based on the species tree (i.e., native copy), but not with the *A. semialata* copy in a phylogenetic position suggestive of LGT (Fig. 4 and *SI Appendix*, Figs. S2 and S5). Fourth, the existence of the 26 remaining candidates was confirmed by independent sequencing runs from multiple samples, ruling out contamination. The 26 candidates passing all of the tests are considered unambiguous LGTs.

A Minimum of Nine Donors Were Involved in the LGTs. For each unambiguous LGT, the putative donor was identified based on the gene trees. Some of the genes could be assigned to precise lineages, such as the *Cenchrus* genus within Cenchrinae and *T. triandra* in Andropogoneae (Figs. 3 and 5 and Dataset S2). Because these two groups were suggested as potential donors of previously reported LGTs in *Alloteropsis* (32), they were purposely included in the genomic datasets. In the case of *Themeda*, the inclusion of multiple accessions further identifies the genomic origin of the donor (Australian *T. triandra* as opposed to African populations; Fig. 3). When the donor species is not included in one of the available genomic datasets, it cannot be identified to the species level, but can be assigned to one of the higher-level groups supported by the coalescence species tree (*SI Appendix*, Fig. S4), with an accuracy that depends on the

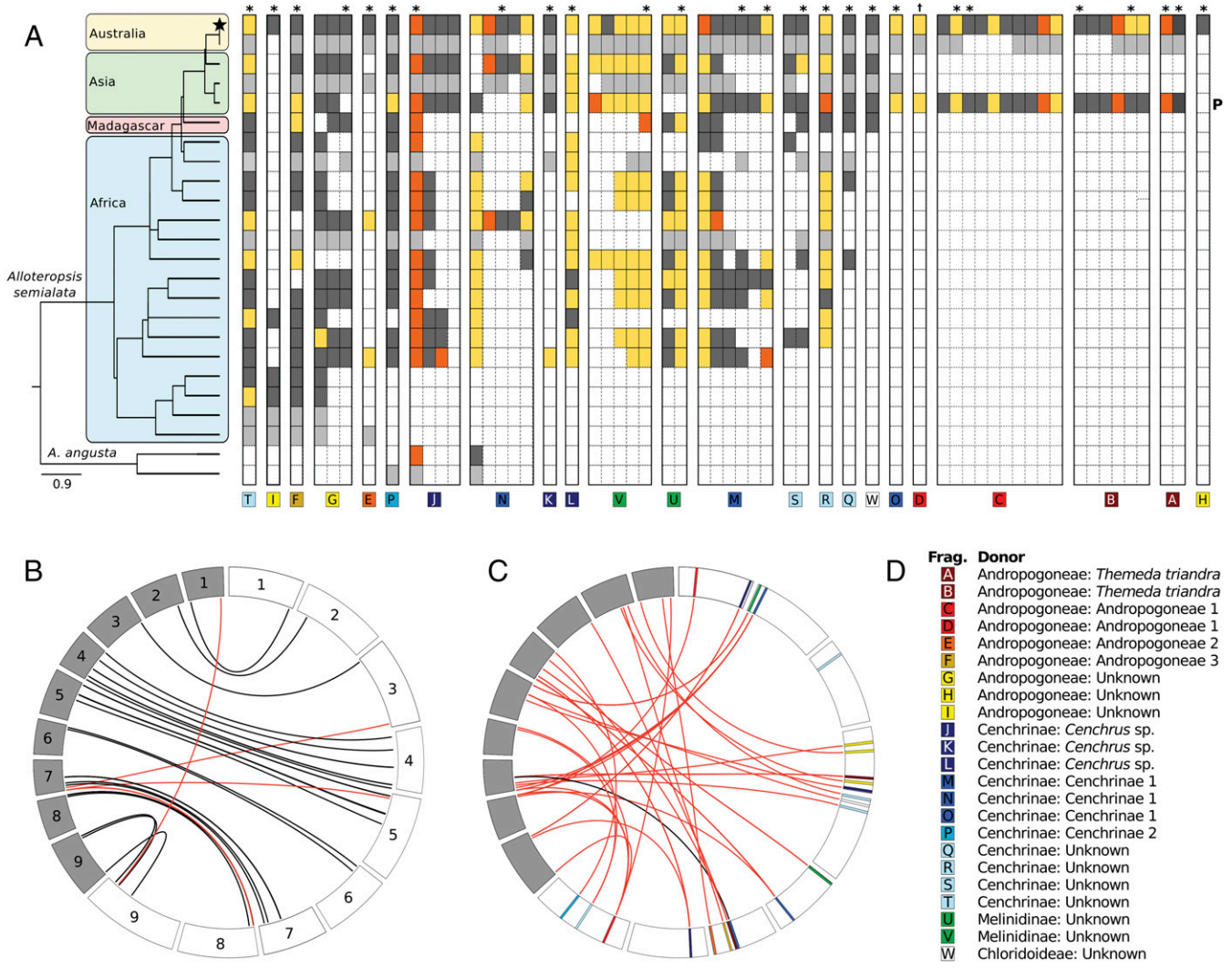

Fig. 4. Phylogenetic and genomic distribution of LGTs in *Alloteropsis*. (*A*) The distribution of the 59 laterally acquired genes (primary and secondary) across the coalescence phylogenetic tree of *Alloteropsis* (extracted from *SI Appendix*, Fig. S4) is shown. Each tip represents a different population, with the reference genome denoted with a star, and the Philippines population indicated with a boldface P on the *Right*. Boxes on the phylogeny delimit geographic origins within *A. semialata*. The LGTs are organized into the 23 acquired genomic fragments, which are labeled at the *Bottom* and the primary candidates within each fragment are indicated by an asterisk (†, gene duplicate from fragment C). The fragments are sorted by the approximate order of their acquisition, with those spread most widely across *A. semialata* phylogeny on the *Left*. Within each fragment, genes are ordered based on their position in the *A. semialata* reference genome. For populations with no corresponding expression data, only presence is shown (light gray). For populations with matched expression data, LGTs present but not expressed are shown in dark gray; those expressed in yellow show ≥1 rpkm; and those expressed at a higher level than the native ortholog are shown in orange (see Dataset S1 for details). Gene presence was inferred based on the number of reads mapping to the coding region in the reference genome (see Dataset S1 for details). The positions of (*B*) native orthologs and the (*C*) LGTs detected in the genome of *A. semialata* (in white) are compared with that of orthologs from *S. italica* (gray). The nine chromosomes are numbered (1–9) and oriented based on their synteny. Genes in syntenic positions are connected by black lines, while those in distinct genomic locations are connected by red lines (see *SI Appendix*, Fig. S2 for details of synteny analyses). (*D*) The donors of the 23 LGT fragments, shown with colors and letters as in *A* and *C*, are listed with the main group and subgroup, as in Fig. 2 and in *SI Appendix*, Fig. S4.

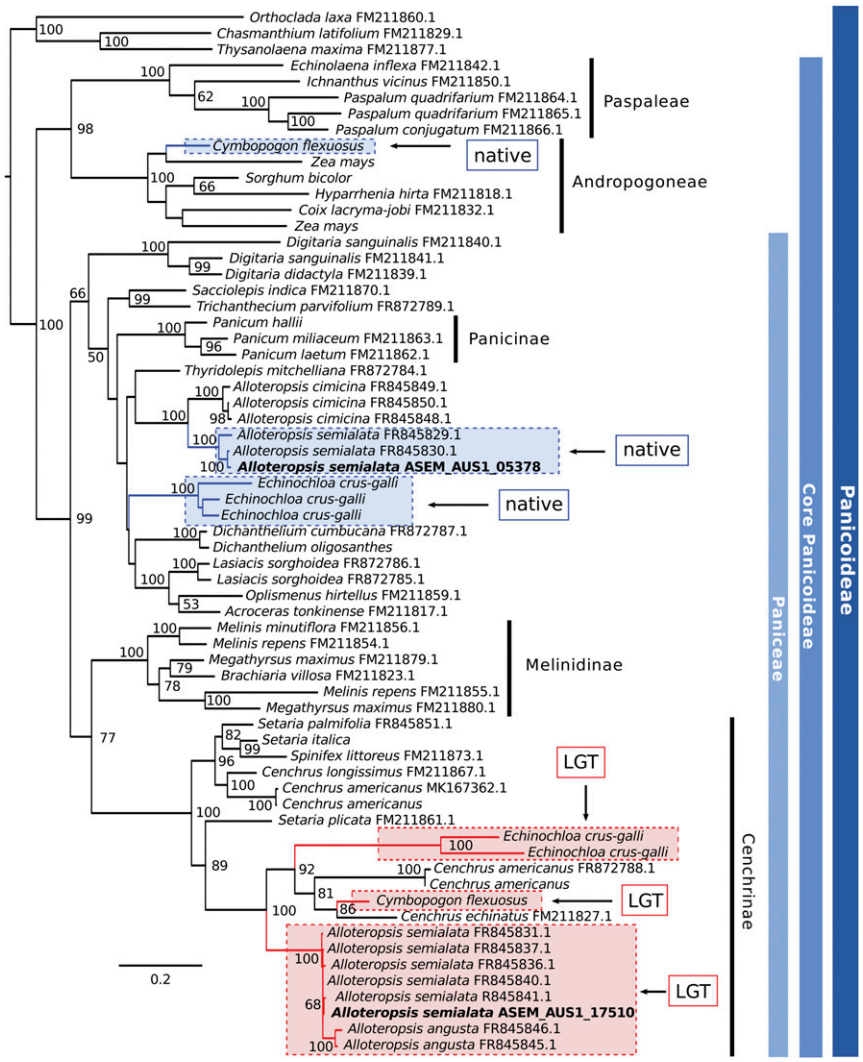

Fig. 5. Phylogenetic tree of Panicoideae genes encoding phosphoenolpyruvate carboxykinase (PCK). The maximum likelihood phylogeny was inferred from gene regions extending from exon 2 to exon 10, and including introns. The three LGTs are highlighted in red, and their corresponding native copies are highlighted in blue. Sequences were either extracted from complete genomes (or transcriptome for *Cymbopogon flexuosus*) or retrieved from GenBank (accession nos. shown). Bootstrap supports ≥50% are indicated near nodes and the *A. semialata* reference genome sequences are shown in bold. Branch lengths are given in expected substitutions per site. The main groups are delimited on the *Right*, as in Fig. 2.

resolution of the gene tree and the sampling density in the clade containing the donor. It is therefore possible that events associated with the same higher-level group correspond to different species, which would increase the number of donors. Taking into account only the nonoverlapping groups, the phylogenetic positioning of the unambiguous candidates indicates at least nine different donors from the Andropogoneae, Cenchrinae, Melinidinae, and Chloridoideae groups (Fig. 2 and *SI Appendix*, Fig. S4). These donors diverged from *Alloteropsis* between 20 and 40 My ago and are separated by multiple speciation events that generated thousands of descending species (32, 47, 48).

The Genes Were Transferred as Part of Large Genomic Blocks. The locations of the 26 unambiguous LGTs in the reference genome were determined, and the amount of DNA involved in the transfers was assessed by mapping reads of close relatives and putative donors onto the reference genome (Fig. 6, *SI Appendix*, *Supplementary Methods*, and Dataset S4). These 26 primary LGTs are located on 23 different fragments of foreign DNA distributed throughout the reference genome, including multiple chromosomes (Fig. 4). We identified protein-coding genes surrounding these primary LGTs and inferred phylogentic trees for them. In total, 26 of the phylogenetic trees built for the surrounding genes supported an LGT scenario involving the same donor as the adjacent primary candidate (*SI Appendix*, Table S3 and Dataset S5). For a further seven genes surrounding the primary candidates, homologs were present in an insufficient number of species or the gene was truncated and too short to infer well-supported phylogenetic trees. We therefore used a combination of mapping of reads from relatives of the putative donor (Dataset S4), BLAST searches, and synteny analyses to support their LGT origin (*SI Appendix*, Table S3), bringing the total of secondary candidates to 33 (*SI Appendix*, *Supplementary Methods*). Seven of these 33 candidates had been detected by our pipeline, but discarded in the second filter because of a lack of resolution of the trees or low statistical support. The identity of the donor was refined, taking into account all genes in each fragment.

Besides multiple protein-coding genes, some of the identified LGT fragments contain long stretches of noncoding DNA with identity to the putative donors above 90% (approximate cutoff in identity for read mapping; Fig. 6 and Dataset S4). In the case of recent LGTs from donors closely related to those included in our dataset, the fragments of laterally acquired DNA could be delimited with confidence and were up to 169,972 bp long with >99% identical coding regions (e.g., for the two genes in LGT fragment A), and highly similar intergenic regions (e.g., 97.2% identical over 45.7 kb in fragment A; Fig. 6). Identifying laterally acquired noncoding DNA is difficult for fragments acquired a long time ago or for which close relatives of the donor have not been sampled, but some of them could be much larger (e.g., fragment N in Fig. 6), while others are limited to one

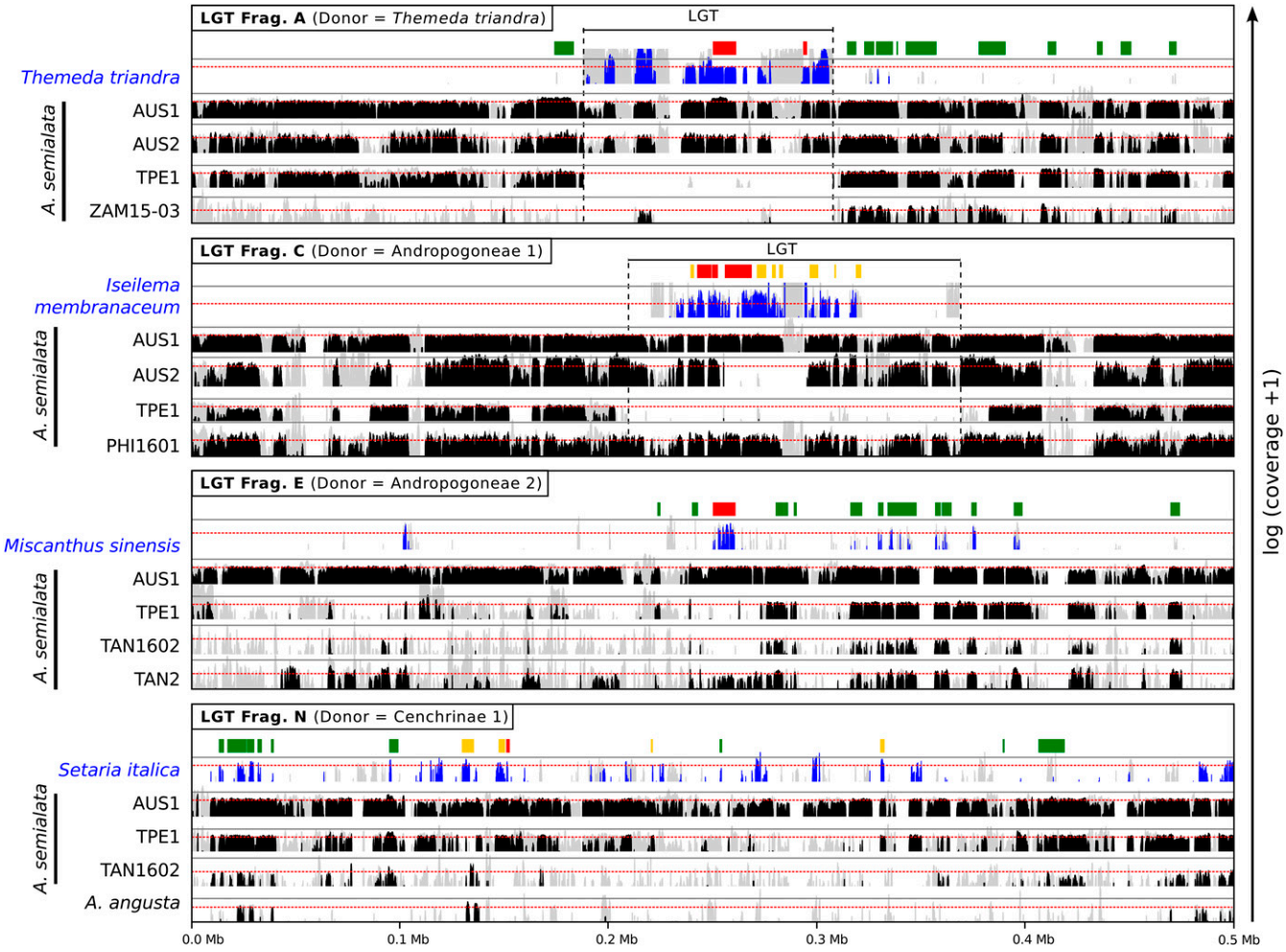

Fig. 6. Genomic context of LGTs in the *A. semialata* reference genome. For four LGT fragments, a 0.5-Mb genomic region is shown with high-confidence protein-coding genes indicated at the *Top* of each panel, in red for primary LGT candidates, orange for secondary LGT candidates, and green for native genes (see Fig. 1 for definitions of primary and secondary LGT candidates). For each fragment, the mapping coverage is shown for the closest relative to the donor in the dataset (in blue), the reference genome AUS1, and three conspecifics or congeners with the three-letter prefix for the *A. semialata* identifiers based on their country of origin. The coverage is shown on a logarithmic scale, with the dotted red lines indicating the coverage expected for single-copy DNA and the gray lines for the coverage expected for a five-copy DNA segment. Black/blue bars represent mapping quality ≥20, while gray bars have mapping quality <20 and include reads that map in multiple locations, indicative of repeats. Valid read alignments have a nucleotide identity of ≥90%. All read lengths are 250 bp except for *Iseilema membranaceum* (151 bp), *Miscanthus sinensis* (100 bp), and *S. italica* (95 bp). The size of the laterally acquired region is indicated by a black bar at the *Top* for fragments A and C, but its delimitation is ambiguous for fragments E and N. See Dataset S4 for details.

protein-coding gene and at most a small amount of flanking DNA (fragment E in Fig. 6).

The number of protein-coding genes contained within the laterally acquired fragments ranged from 1 (13 fragments) to 10 (in fragment C) (Fig. 4). When multiple genes were identified in a laterally acquired fragment, the genomic segment shows evidence of synteny with the genomes of the two other Panicoideae investigated (Fig. 4 and *SI Appendix*, Fig. S2 and Table S3). We stress, however, that the native and laterally acquired genes of *A. semialata* are always found in different parts of the genome, with almost all native copies being syntenic with the position of the ortholog in the *S. italica* and *S. bicolor* genomes (Fig. 4 and *SI Appendix*, Figs. S2 and S5). By contrast, the LGT are not syntenic with orthologs from *S. italica* and *S. bicolor*, with one exception (fragment O; Fig. 4 and *SI Appendix*, Figs. S2 and S5).

Transposable Elements Were also Transferred and Later Duplicated. We assembled a partial genome of *T. triandra* and identified *T. triandra* and *A. semialata* TEs in their respective genome assemblies (*SI Appendix*, *Supplementary Methods*). In total, 186,270 TEs were annotated in the *A. semialata* reference genome, accounting for 51% of the assembly (0.39 Gb; *SI Appendix*, Table S4). As expected, they appear prevalent near centromeres (*SI Appendix*, Fig. S3). The TE sequences were then clustered and phylogenetic trees inferred. Combined with coverage analyses, these led to the identification of 92 TEs acquired from *T. triandra* in the *A. semialata* genome (*SI Appendix*, *Supplementary Methods*). Several of these TEs are located within the two fragments acquired from *T. triandra* containing protein-coding genes and potentially have daughter copies located in other genomic locations (*SI Appendix*, Fig. S3).

LGTs Happened at Different Times During *Alloteropsis* Diversification. Using resequencing (~10× coverage) and high-coverage (~40× coverage) data for an additional 24 *Alloteropsis* populations representing the genetic diversity in this species (*SI Appendix*, Fig. S4), we were able to establish the distribution of the LGTs across the *Alloteropsis* phylogenetic tree, thereby estimating the relative timing of the transfers (*SI Appendix*, *Supplementary Methods*). Fragment H was unique to the Australian reference

genome, and four other fragments (A–D) are restricted to the other Australian accession (AUS2) as well as the accession from the Philippines (PHI1), indicating recent acquisitions. The other LGTs were also observed in a number of individuals from *A. semialata* and/or the sister species *Alloteropsis angusta* (Fig. 4). Some are present in most *A. semialata* accessions (T, I, F, and G), suggesting an early acquisition followed by retention of the genes. Others are distributed among more distantly related accessions, suggesting early acquisitions followed by losses or introgression among *A. semialata* populations after the transfer (Fig. 4).

LGTs Added Functional Diversity to the Recipient Genomes. RNA-Seq data for 16 populations of *A. semialata* and one of the sister species *A. angusta* were mapped to the coding sequences from the *A. semialata* reference genome, and the expression levels of LGTs and their native orthologs were estimated in leaf and root samples (*SI Appendix*, *Supplementary Methods*). Over 59% (35 out of 59) of the primary and secondary LGTs were expressed [>1 reads per kilobase of transcript per million mapped reads (rpkm)] in at least one population (Fig. 4 and Dataset S1). The mean expression level of 12 laterally acquired genes was higher than their respective native homolog in at least one *A. semialata* population (Fig. 4 and Dataset S1), and in one example the expression of the native ortholog seems to have been replaced by the LGT (ASEM_AUS1_12633; Dataset S1). The functions of the laterally acquired genes include those known to be involved in C_4 photosynthesis as well as loci associated with disease resistance and abiotic stress tolerance (*SI Appendix*, Table S5).

LGTs Involving Other Recipient Grasses Are Detected. Our phylogenetic trees suggest LGTs involving recipients other than *A. semialata*. Indeed, the visual inspection of the gene trees for primary LGTs including the 147 species identified 10 genes with bootstrap support for positions suggestive of LGT in grasses other than the reference genome (*SI Appendix*, Table S6 and Dataset S2). For three of them, the lack of sequencing replicates for these samples generated in other projects meant that contaminations could not be ruled out (*SI Appendix*, Table S6). However, seven genes from individuals other than the reference genome were statistically supported in a position suggestive of LGT by AU tests (Bonferroni-corrected P value <0.05; *SI Appendix*, Table S7), and this conclusion was supported by replicate datasets (*SI Appendix*, Table S6). This included two genes encoding a C_4 enzyme that had been previously identified in some non-Australian populations of *A. semialata* (Dataset S2) (32, 36). In addition we identified LGTs in five other Panicoideae genomes (*SI Appendix*, Table S6).

For two genes containing non-*A. semialata* LGTs we had sufficient genome data for the recipient species to assemble full-length gene sequences. These data were supplemented with similar full-length sequences from published genomes and the NCBI nucleotide database to infer gene trees using intron and exon sequences. This confirmed that the gene encoding phosphoenolpyruvate carbokykinase (PCK; ASEM_AUS1_17510 on fragment L), an enzyme involved in some subtypes of C_4 photosynthesis (49), was laterally acquired by *A. semialata*, *Echinochloa*, and *Cymbopogon* (Fig. 5). The three LGTs of PCK, supported by multiple datasets, did not form a monophyletic group, suggesting they result from independent transfer events (Fig. 5). Transcriptome datasets for these species indicate that both *Cymbopogon* and *Echinochloa* express the LGTs in their leaves, where they likely play a role in their C_4 photosynthetic pathway (50, 51). Despite the smaller sample size, the phylogenetic tree inferred from sequences homologous to ASEM_AUS1_20550 similarly confirmed the LGT scenario for this gene observed in *Alloteropsis cimicina* (*SI Appendix*, Fig. S6). For ASEM_AUS1_20550 the donor for the LGT detected in the reference genome belonged to the Andropogoneae, whereas the gene detected in *A. cimicina* was acquired from a Cenchrinae (*SI Appendix*, Fig. S6). The native copy of *A. cimicina* was also detected, and transcriptome data show that the LGT is expressed at a higher level than the native copy.

Discussion

Multiple Laterally Acquired Genes in the Genome of *A. semialata*. Using a combination of stringent phylogenetic and genomic analyses (Fig. 1), we identify 59 genes in the genome of the grass *A. semialata* that were laterally acquired from other grasses. Our pipeline was designed to rule out the three main alternative explanations to LGT, namely: (*i*) unrecognized parology by comparing synteny with other grass genomes, (*ii*) contamination by having independent sequencing supporting the existence of the genes, and (*iii*) convergent evolution by using different data partitions. In addition, long reads spanning the laterally acquired and native DNA prove that the fragments are integrated in the genome of *A. semialata* (*SI Appendix*, Fig. S7). The LGTs are moreover supported a posteriori by our genomic analyses, which show that half of the primary candidates are flanked by coding, and in some cases noncoding, regions with a high similarity to the same putative donors (Fig. 6 and Dataset S4). The 10 fragments that contain multiple protein-coding genes represent the most unequivocal cases of LGTs and demonstrate that large stretches of DNA containing numerous genes can be transferred among distantly related grass species.

The number of LGTs reported here is likely an underestimate. For example, LGTs that lack phylogenetic informativeness because the genes are too short, or are not present in a sufficient number of grass species, would be excluded by our analysis. More importantly, we focused on gene transfers among distant relatives to differentiate LGTs from other processes that can create discordance between gene and species trees, such as incomplete lineage sorting and hybridization (Fig. 2). Similarly, we focused on relatively recent LGTs that have accumulated during the diversification of the *Alloteropsis* genus, because ancient LGT would alter deep branching patterns that can be detected only when the donor and recipient are extremely distant [e.g., ferns and mosses (33)]. As a direct consequence of the methodology, all transfers involving donors that are part of poorly resolved clades within the Paniceae tribe (Fig. 2) or transfers that happened before the diversification of *Alloteropsis* would remain undetected. In addition, our power to detect LGTs depends directly on the availability of genomic data for close relatives of the donor, particularly when it comes to accurately determining the size of the acquired DNA fragment (Fig. 6). Thus, the 59 LGTs reported here could be only a subset of those existing in the genome of *A. semialata*.

In total, the detected LGTs belong to 23 genomic DNA fragments (Fig. 4). These fragments were laterally acquired from at least nine different grass donors, although the number might be higher if grasses from the same lineage independently provided LGTs (Fig. 4). Using genomic datasets from multiple *Alloteropsis* populations allowed us to establish the distribution of each of these fragments within the species. Inferring the presence of a gene from resequencing data can be problematic if the gene is truncated, or located in regions of the genome with reduced sequencing depth. Based on our estimates, fragment H is unique to the reference genome and likely represents the most recent acquisition. Several fragments (A–D) were likely acquired around the time *A. semialata* colonized Australia, as they are restricted to Australian and Filipino accessions, with the latter probably a result of recent admixture from Australia (Fig. 4). Other LGT fragments are shared by a majority of *A. semialata* accessions and were likely acquired near the origin of this species ~2 My ago (e.g., fragments T and F; Fig. 4) (36, 41). In some cases, the patchy distribution of LGT fragments could be due to secondary losses after ancient acquisitions, as supported by observed genetic variation among the different *A. semialata* populations

that has likely accumulated since the LGT was acquired (e.g., fragment E; Dataset S2). In other cases low levels of genetic variation among accessions of *A. semialata* and *A. angusta* suggest that the patchy distribution results from more recent acquisition followed by introgression into different populations (e.g., fragments K and N; Figs. 4 and 6 and Datasets S2 and S5) (52). Overall, the evidence of fragments being acquired at different points suggests that the diversification of *Alloteropsis* has been punctuated by repeated bouts of LGT (Fig. 4*A*).

Transfers of Large DNA Blocks Spread Functional Genes. The 23 laterally acquired fragments are widely distributed across the genome of *A. semialata* (Fig. 4*C*). While divergence of noncoding DNA hampers a precise delimitation of the acquired fragments in more ancient LGTs, or those for which genome data of a close relative of the donor are missing (e.g., fragments E and N in Fig. 6), recent LGTs with a sampled donor can be shown to be at least 170 kb long and are composed of genic as well as noncoding regions (e.g., fragments A and C in Fig. 6). In the case of two fragments acquired from *T. triandra* (A and B), TE phylogenetic trees and inferences of their recent activity show that some TEs, which were acquired as part of the large DNA fragments, have subsequently transposed to new regions (*SI Appendix*, Fig. S3*B*). The laterally acquired fragments also have TEs specific to *Alloteropsis*, which were likely inserted after the acquisition, highlighting rearrangements between the native and foreign parts of the genome. Other laterally acquired TEs were detected around the genome outside of the large block of DNA and might represent TEs that escaped from large DNA blocks or elements that were acquired on their own.

Genomic rearrangements that happened after the transfer are also visible in the gene content of the laterally acquired fragments. Erosion is evidenced by gene loss in some accessions (e.g., part of fragment C in one of the Australian samples; Figs. 4 and 6), as well as pseudogenizing mutations in others (*SI Appendix*, Fig. S8). However, 59% of the laterally acquired genes are expressed in at least one population under the conditions evaluated (Fig. 4). In all cases, the donor and the recipient evolved independently for tens of millions of years before the transfers, leaving ample time for adaptive evolution and functional diversification. The genetic exchanges therefore potentially brought in genes for novel attributes. Eight of the fragments contain at least one gene that is either novel, has replaced the function of the native copy that became a pseudogene, or is expressed at a higher level than the native copy (Fig. 4 and Dataset S1). The functional LGTs have therefore added novelty to the genetic apparatus of *Alloteropsis*, including a variety of disease resistance and abiotic stress response loci (*SI Appendix*, Table S5), in addition to the previously reported photosynthetic genes (32). The advantage of each of the laterally acquired genes is yet to be determined by targeted functional studies, but selection for the novel functions is likely responsible for the retention of LGTs through time (Fig. 4). Genes that underwent adaptive shifts in some subgroups of grasses might be especially likely to be retained after transfer to other groups, potentially contributing to the three independent LGTs for the C_4 enzyme PCK (Fig. 5).

Different Processes Might Have Transferred the Genes. The mechanisms underlying the reported grass-to-grass LGTs remain elusive and might vary between events. The nonsyntenic localization of LGTs and their coexistence with native copies argues against classical introgression involving sexual reproduction and chromosomal recombination. The movement of genes could have occurred among genomes following the transient cohabitation of chromosomes from different species within the same nucleus, for example in allopolyploids that can subsequently backcross with diploids (53). In grasses, the growth of pollen tubes on the stigma of distant relatives can be used experimentally to trigger embryo development with only occasional transfer of paternal DNA (54, 55), providing a more likely mechanism for chromosomal exchanges. Occasional interspecific cell-to-cell contact could also occur through root-to-root interactions between grasses growing in multispecies clumps as observed in savannas or among pollen tubes of different species growing on the same stigma. Cell-to-cell contacts are known to allow movement of DNA across distinct nuclei in grafts (56) or host/parasite interactions (22–30). Since *A. semialata* can propagate vegetatively via rhizomes, both transfers into the seeds and into parts of the root system would allow the long-term integration of LGTs into the germline. Independently of the exact mechanism, our genomic investigations show that genes are recurrently passed among distant species of grasses, and provide a novel source of genetic variation for selection to act upon.

While we screened for nonangiosperm donors, all of the detected LGTs came from grasses. This bias in the donor identity might represent some genomic or physical incompatibilities (e.g., lack of wind dispersal of pollen or different root architecture) that limited exchanges with nongrass angiosperms. All grass donors were from the subfamily Panicoideae, with one exception. Within Panicoideae, two groups (Andropogoneae and Cenchrinae) have contributed the vast majority of fragments, while other groups with similar genomic resources (e.g., Panicinae and Paspaleae; Fig. 2 and *SI Appendix*, Table S1) were not involved in any of the detected LGTs. This bias can largely be explained by geographic patterns, as both Andropogoneae and Cenchrinae species frequently co-occur with *A. semialata* in large populations throughout Africa, Asia, and Australia, whereas in contrast, Paspaleae species are mainly found in South America where *A. semialata* does not occur. Whether other groups (e.g., Panicinae) had opportunities to exchange genes with *Alloteropsis* remains to be formally assessed and may ultimately contribute to identifying the properties that promote LGT.

Conclusions

Using genomic tools and stringent phylogenetic criteria, we have shown that the genome of an Australian *A. semialata* individual contains at least 59 genes laterally acquired from a minimum of nine different donors (Figs. 2 and 4). Large-scale pollen dispersal and vegetative growth, which might facilitate cell-to-cell contacts and subsequent LGTs, are widespread among perennial grasses, and LGT is therefore likely to be frequent in this group. Transfer of specific segments of noncoding DNA among members of the grass family has previously been reported (24, 34, 35), but the widespread transfer of functional genes documented here shows that this process is likely to have consequences for adaptation. We also detect functional LGTs among grass species other than *Alloteropsis*, with recipients in the genera *Cymbopogon*, *Danthoniopsis*, *Echinochloa*, and *Oplismenus* (Fig. 5, *SI Appendix*, Fig. S6 and Table S6, and Dataset S2). While these other instances need to be investigated with dedicated genomic work, they do show that *A. semialata* is not exceptional in this group of eukaryotes that might exhibit something approaching a type of pangenome. In particular, future efforts should determine whether the process is pervasive in the family, or whether it is restricted to certain growth forms or ecological types. The evidence presented here already shows that the widespread transfer of functional genetic elements reported for *Alloteropsis* might be just the tip of the iceberg, with functional LGTs among grasses, and potentially other groups of plants, having remained undetected because of limited taxon sampling and a lack of dedicated searches. We conclude that LGT might constitute an underappreciated contributor to the functional diversification of some groups of plants, which can act as a source of genetic variation, potentially of adaptive significance.

Materials and Methods

This section gives a summary of the extensive methodology, which is detailed in *SI Appendix, Supplementary Methods*. In short, a chromosome-level

reference genome was generated for a single Australian plant of *A. semialata*, selected because it was previously shown to contain a gene laterally acquired from a member of the *Themeda* genus (32, 36). Ab initio gene prediction was used to annotate the reference genome using a combination of *A. semialata* transcriptome data and protein sequences from model Panicoideae species (*S. italica* and *S. bicolor*). A combination of similarity and phylogenetic analyses were used to identify unambiguous LGT among these genes (Fig. 1).

We first used high-coverage Illumina genome data (~40 Gbp per sample) to scan the Australian *A. semialata* genome for loci that are more similar to a potential donor species than to a close relative, representing LGT candidates (Fig. 1). Datasets for three close relatives were used to capture LGTs that occurred at different time points during the diversification of *Alloteropsis*, and equivalent data were generated or retrieved from the literature for 17 other species, including *Themeda*, representing potential donors distributed across the grass family.

All candidates from the initial genome-wide scan were subsequently verified using phylogenetic trees. Up to 147 grass species were included and genes were considered to be laterally acquired if they fulfilled a number of stringent criteria (Fig. 1). The genomic fragments containing the detected LGTs were characterized, and the potential adaptive significance of the LGTs was assessed using RNA-Seq data and functional annotation. Finally, different classes of TEs were annotated using existing pipelines, and phylogenetic trees combined with coverage analyses were used to identify those TEs transferred from *T. triandra* to *A. semialata*.

ACKNOWLEDGMENTS. We thank Tricia Handasyde (Department of Parks and Wildlife, Western Australia) for providing samples. This work was funded by a Natural Environment Research Council (NERC) Grant (NE/M00208X/1), an European Research Council Grant (ERC-2014-STG-638333), and a Royal Society Research Grant (RG130448). P.-A.C. and P.N. were supported by Royal Society University Research Fellowships (URF120119 and URF138573, respectively). Library preparation and sequencing were carried out by Edinburgh Genomics, University of Edinburgh, Dovetail Genomics, and by the Georgia Genomics and Bioinformatics Core, University of Georgia. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), Medical Research Council (MR/K001744/1), and Biotechnology and Biological Sciences Research Council (BB/J004243/1). G.B. is a member of the Laboratoire Evolution & Diversité Biologique part of the Laboratoires d'Excellence entitled TULIP and CEBA managed by Agence Nationale de la Recherche (ANR-10-LABX-0041 and ANR-10-LABX-25-01).

1. Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends Ecol Evol* 23:38–44.
2. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513–518.
3. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605–618.
4. Boothby TC, et al. (2015) Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci USA* 112:15976–15981.
5. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol* 16:50.
6. Koutsovoulos G, et al. (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci USA* 113: 5053–5058.
7. Martin WF (2017) Too much eukaryote LGT. *BioEssays* 39:1700115.
8. Martin WF (2018) Eukaryote lateral gene transfer is Lamarckian. *Nat Ecol Evol* 2:754.
9. Salzberg SL (2017) Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol* 18:85.
10. Roger AJ (2018) Reply to 'Eukaryote lateral gene transfer is Lamarckian'. *Nat Ecol Evol* 2:755.
11. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.
12. Savory F, Leonard G, Richards TA (2015) The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog* 11:e1004805.
13. Husnik F, McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol* 16:67–79.
14. Peccoud J, Loiseau V, Cordaux R, Gilbert C (2017) Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci USA* 114:4721–4726.
15. Savory FR, Milner DS, Miles DC, Richards TA (2018) Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol Biol Evol* 35:1887–1900.
16. Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328:624–627.
17. Grbić M, et al. (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479:487–492.
18. Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197–201.
19. Won H, Renner SS (2003) Horizontal gene transfer from flowering plants to *Gnetum*. *Proc Natl Acad Sci USA* 100:10824–10829.
20. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci USA* 101:17747–17752.
21. Rice DW, et al. (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342:1468–1473.
22. Davis CC, Wurdack KJ (2004) Host-to-parasite gene transfer in flowering plants: Phylogenetic evidence from Malpighiales. *Science* 305:676–678.
23. Mower JP, Stefanović S, Young GJ, Palmer JD (2004) Plant genetics: Gene transfer from parasitic to host plants. *Nature* 432:165–166.
24. Diao X, Freeling M, Lisch D (2006) Horizontal transfer of a plant transposon. *PLoS Biol* 4:e5.
25. Yoshida S, Maruyama S, Nozaki H, Shirasu K (2010) Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328:1128.
26. Xi Z, et al. (2012) Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* 13:227.
27. Zhang Y, et al. (2013) Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant *Phelipanche aegyptiaca* and related species. *BMC Evol Biol* 13:48.
28. Park S, et al. (2015) Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New Phytol* 208:570–583.
29. Yang Z, et al. (2016) Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proc Natl Acad Sci USA* 113: E7010–E7019.
30. Skippington E, Barkman TJ, Rice DW, Palmer JD (2017) Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. *BMC Plant Biol* 17:49.
31. Vallenback P, Jaarola M, Ghatnekar L, Bengtsson BO (2008) Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. *Mol Phylogenet Evol* 46:890–896.
32. Christin PA, et al. (2012) Adaptive evolution of $C_{(4)}$ photosynthesis through recurrent lateral gene transfer. *Curr Biol* 22:445–449.
33. Li FW, et al. (2014) Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc Natl Acad Sci USA* 111:6672–6677.
34. Mahelka V, et al. (2017) Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc Natl Acad Sci USA* 114:1726–1731.
35. El Baidouri M, et al. (2014) Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res* 24:831–838.
36. Olofsson JK, et al. (2016) Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Mol Ecol* 25:6107–6123.
37. Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14:860–869.
38. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316:1862–1866.
39. Ammiraju JSS, et al. (2008) Dynamic evolution of oryza genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* 20:3191–3209.
40. Westwood JH, Yoder JI, Timko MP, dePamphilis CW (2010) The evolution of parasitism in plants. *Trends Plant Sci* 15:227–235.
41. Lundgren MR, et al. (2015) Photosynthetic innovation broadens the niche within a single species. *Ecol Lett* 18:1021–1029.
42. Bennetzen JL, et al. (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol* 30:555–561.
43. Paterson AH, et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556.
44. Bennetzen JL, Freeling M (1997) The unified grass genome: Synergy in synteny. *Genome Res* 7:301–306.
45. Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci USA* 95:1971–1974.
46. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
47. Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol* 193:304–312.
48. Soreng RJ, et al. (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae). *J Syst Evol* 53:117–137.
49. Hatch MD, Kagawa T, Craig S (1975) Subdivision of C_4-pathway species based on differing C_4 acid decarboxylating systems and ultrastructural features. *Funct Plant Biol* 2:111–128.
50. Meena S, et al. (2016) *De novo* sequencing and analysis of lemongrass transcriptome provide first insights into the essential oil biosynthesis of aromatic grasses. *Front Plant Sci* 7:1129.
51. Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA (2018) Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Mol Biol Evol* 35:94–106.
52. Dunning LT, et al. (2017) Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* 71: 1541–1555.
53. Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst* 29:467–501.
54. Laurie DA, Bennett MD (1986) Wheat × maize hybridization. *Can J Genet Cytol* 28: 313–316.
55. Riera-Lizarazu O, Rines HW, Phillips RL (1996) Cytological and molecular characterization of oat x maize partial hybrids. *Theor Appl Genet* 93:123–135.
56. Stegemann S, Bock R (2009) Exchange of genetic material between cells in plant tissue grafts. *Science* 324:649–651.